ARMY RESEARCH LABORATORY

Analysis of Wavelet Transform Multiscale Products for Step Detection and Estimation

by Brian M. Sadler and Ananthram Swami

ARL-TR-1664 August 1998

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory
Adelphi, MD 20783-1197

ARL-TR-1664 **August 1998**

Analysis of Wavelet Transform Multiscale Products for Step Detection and Estimation

Brian M. Sadler and Ananthram Swami
Information Science and Technology Directorate

Abstract

We consider discrete wavelet transform (DWT) multiscale products for detection and estimation of steps. Here the DWT is an overcomplete approximation to smoothed gradient estimation, with smoothing varied over dyadic scale, as developed by Mallat and Zhong. We show that the multiscale product approach, as first proposed by Rosenfeld for edge detection, is a nonlinear whitening transformation. We characterize the resulting non-Gaussian heavy-tailed densities. The results may be applied to edge detection with a false alarm constraint. The response to impulses, steps, and pulses is also characterized. A general closed-form expression for the Cramer-Rao bound (CRB) for discrete and continuous-time step-change location estimation in independent identically distributed non-Gaussian noise is developed, generalizing previous results. We consider location estimation using multiscale products, and compare results to the appropriate CRB.

Contents

Figures

Tables

1. Introduction

Step (or edge) detection and estimation is a fundamental problem in many areas of signal and image processing that involves basic statistical tradeoffs. A classical approach to this problem is based on gradient estimation, noting that the gradient will be large when a noise-free step is encountered. This approach has been heavily exploited in image processing; e.g., see Jain [13, chapter 9]. Many of the basic image edge-detection operators (see Roberts, Prewitt, Sobel) reduce in 1-D to a finite impulse response (FIR) filter with response [–1, 0, 1]. More general extensions, so-called filtered derivative methods, combine smoothing with gradient estimation to reduce noise effects. These methods are attractive for their low-complexity linear implementations. In addition, they tend to be localized, and are therefore useful in cases such as images that may be characterized by relatively high signal-to-noise ratio (SNR) and multiple closely spaced change points.

The filtered derivative approach is motivated by application of linear filters whose output is maximized when a step is encountered, while attempting to maintain good location estimation (localization) and low probability of false alarm. A filtered derivative method that has received a lot of attention is the derivative of Gaussian (dG), which estimates the gradient after smoothing with a Gaussian function. The dG approach can be derived under criteria of detection and localization (see Canny [7], Tagare and deFigueiredo [32], and Koplowitz and Greco [15]). The problem can also be formulated in terms of zero crossings of the second derivative, such as the Laplacian of Gaussian approach, which is equivalent to dG in 1-D. Attempting to achieve simultaneous detection and estimation results in a tradeoff between the level of smoothing and the variance of the estimated step location, and this tradeoff is sensitive to the edge shape and SNR. It is therefore of interest to study gradient estimation techniques that exploit multiple levels of smoothing (multiscale) simultaneously.

Many alternative techniques are based on formulating the problem as a change in distribution, such as a step change in the mean. These methods often assume a moderate to large sample size around a *single* point of change, and find application in such (typically 1-D) problems as fault detection or vibration monitoring; e.g., see Basseville and Nikiforov [1] and references therein. This problem may be formulated as one of quickest change detection in an on-line setting; a Bayes-optimal approach was formulated by Shiryaev in 1961 [30]. An important approach in this context is

the CUSUM (cumulative sum) algorithm; e.g., see Basseville and Nikiforov [1, chapter 2]. These methods may be extended to include other changes in statistical description, such as changes in second-order statistics. We note that CUSUM and other higher complexity approaches may outperform filtered derivative methods at low SNR when sufficient data are available (e.g., see Basseville et al. [2–4]), although problems may be encountered for closely spaced change points.

It is well known that wavelets may be used to characterize the local regularity of signals, and may be applied for singularity detection and characterization; e.g., see Mallat and Hwang [21]. In this regard, Mallat and Zhong [22] developed a discrete wavelet transform based on smoothed gradient estimation, where the level of smoothing varies with (dyadic) scale, conveniently placing the earlier work of Canny and others in the wavelet context. This implementation is based on splines, and closely approximates the derivative of Gaussian estimator; hereafter, we refer to the implementation in Mallat and Zhong [22] as the MZ-DWT algorithm (Matlab M-file listings are provided in the appendix for the forward and inverse MZ-DWT algorithm). This approach is nonorthogonal, so as to preserve regularity information at each point in time for each scale. Thus the MZ-DWT is simply composed of a bank of FIR filters that are dyadic in scale, but whose outputs are not downsampled. Each filter corresponds to dG with various levels of smoothing. This has been applied to filtering (denoising) [21,34], compression [22], ECG characterization [18], and acoustic shockwave detection [29].

A central question, then, is how to process and combine the multiscale gradient information. Mallat and Zhong [21] estimated Lipschitz exponents by tracking local maxima across scales, an approach that is rather sensitive to noise. Li et al. [18] developed ad hoc modifications for the ECG problem. We also note the earlier work of Witkin [33], who developed ideas for multiscale characterization of signals. Working on image processing before the advent of wavelets, Rosenfeld and coworkers suggested an interesting idea, to form multiscale point-wise products [26–28]. This approach is intended to enhance multiscale peaks due to edges, while suppressing noise, by exploiting the multiscale correlation due to the presence of the desired signal in a direct (albeit nonlinear) way. It is interesting to note that Rosenfeld's original work used dyadic scales, both for ease of implementation and because this seemed to work as well as any other combination of scales. The multiscale product idea was recently applied for filtering of magnetic resonance images by Xu et al. [34]. However, statistical analysis of the multiscale product technique is lacking in this work.

In this report, we study the multiscale product approach for edge detection and estimation. We characterize it statistically and evaluate performance for detection and estimation of edges. First, we briefly introduce the MZ-DWT and the multiscale product; numerical examples throughout the report are based on the MZ-DWT implementation. Then we develop the second-order statistics of multiscale products, showing the whitening effect of this nonlinear operation. We consider the step and pulse response as a function of step width, quantifying the suppression of impulses in the multiscale product. We then characterize the first-order probability density function (pdf) of multiscale products, showing that they are, in general, heavy-tailed. The resulting pdfs are applied to detection of edges with a false alarm constraint. We consider step location estimation, and give results based on a general closed-form Cramer-Rao bound (CRB) derivation, provided in the appendix. We close with a discussion of our study.

2. A Non-Orthogonal DWT and Multiscale Products

In this section we provide background and a quick overview of the MZ-DWT algorithm. Matlab M-file listings are provided in the appendix. We then introduce the multiscale product and provide motivating examples.

2.1 Multiscale Gradient Estimation

Consider a continuous time wavelet given by $\phi_s(t) = (1/s)\phi(t/s)$, such that $\phi(t)$ meets appropriate criteria to be a wavelet. The continuous wavelet transform (CWT) of $x(t) \in L^2(R)$ is $W_s x(t) = x(t) * \phi_s(t)$, where $*$ denotes convolution. The CWT is invertible via

$$x(t) = C \int_0^\infty \int_{-\infty}^\infty W_s x(\tau)\, \phi_s^*(\tau - t) d\tau \frac{ds}{s}, \tag{1}$$

where C is a constant that depends on $\phi(t)$, and $\phi^*(t)$ denotes complex conjugation. In this report we restrict ourselves to 1-D; full details and 2-D extensions of the CWT are described by Mallat [21,22].

Here we are concerned with wavelets such that $\phi(t) = du(t)/dt$, with $u(t)$ acting as a local average or smoothing function. Now,

$$W_s x(t) = x(t) * \left(s \frac{du_s}{dt} \right)(t) = s \frac{d}{dt}(x * u_s)(t), \tag{2}$$

provided that $\phi(t)$ is a valid wavelet with zero first moment, so that differentiation and integration may commute [22]. Thus, for appropriate choice of $u(t)$, $W_s x(t)$ can be interpreted as a derivative of a local average of $x(t)$ where the degree of smoothing depends on scale s. Of particular interest here is the case of $u(t)$ closely approximating a Gaussian function; the result is derivative estimation at various levels of smoothing.

Mallat and Zhong [22] developed a discrete wavelet transform (DWT) based on a discrete-time approximation to a Gaussian $u(t)$ using a cubic spline. Consequently, $\phi(t)$ is a quadratic spline. Restricting to dyadic scales, the DWT of $x(n)$, $1 \le n \le N$, consists of

$$W_{2^j} x(n), \quad j = 1, 2, \cdots, J-1\,, \tag{3}$$

where $J = \log_2 N$, plus the remaining coarse scale information denoted by $S_J(n)$. This DWT, consisting of $J \times N$ points, is overcomplete (non-orthogonal). This contrasts with the (perhaps more commonly encountered)

orthogonal wavelet transforms where the number of coefficients decreases with scale. The inverse DWT may be readily computed, enabling filtering and reconstruction (see remark 1 at the end of this section).

To emphasize the linear filtering aspect of equation (3) and to simplify notation, let

$$y_j(n) = W_{2^j}x(n) = \sum_k h_j(k)x(n-k) \tag{4}$$

denote the DWT response to $x(n)$ at the jth scale, where $h_j(n)$ is the impulse response (IR) of the jth DWT filter. The MZ-DWT may be expressed in the Fourier domain as follows. The Fourier transform of $\phi(n)$ is [22]

$$\Phi(\omega) = j\omega\left(\frac{\sin\omega/4}{\omega/4}\right)^4. \tag{5}$$

Denoting the Fourier transform of $h_j(n)$ as $H_j(e^{j\omega})$, then

$$H_j(e^{j\omega}) = \begin{cases} G(\omega), & j=1 \\ G(2\omega)H(\omega), & j=2 \\ G(2^{j-1}\omega)H(2^{j-2}\omega)\cdots H(\omega), & j>2 \end{cases} \tag{6}$$

where

$$G(\omega) = 4je^{j\omega/2}\sin(\omega/2), \tag{7}$$

and

$$H(\omega) = e^{j\omega/2}\left(\cos(\omega/2)\right)^3. \tag{8}$$

$G(\omega)$ and $H(\omega)$ give rise to a time-domain recursion in the MZ-DWT filter coefficients [22]. Note that $H_j(e^{j\omega})$ is purely imaginary and hence $h_j(n)$ is anti-symmetric.

Hereafter we restrict our attention to Mallat and Zhong's implementation (MZ-DWT), although our results are general for a family of linear derivative estimation filters. Here, each filter is an approximation to dG, with smoothing increasing with scale. The IR's for the first five scales of the MZ-DWT are shown in figure 1 (a) through (e), and some frequency response magnitudes of equation (6) are shown in figure 2. Note that the linear regions around the origin in figure 2 correspond to the frequency response of a differentiation filter.

2.2 Multiscale Products

Working before the advent of the wavelet framework, Rosenfeld and coworkers suggested forming multiscale point-wise products [26–28]. This is intended to enhance multiscale peaks due to edges, while suppressing noise,

Figure 1. Impulse responses of MZ-DWT for (a) through (e) first five scales, scale $s = 2^j$, $1 \leq j \leq 5$; (f) multiscale product $p_2(n)$; and (g) $p_3(n)$.

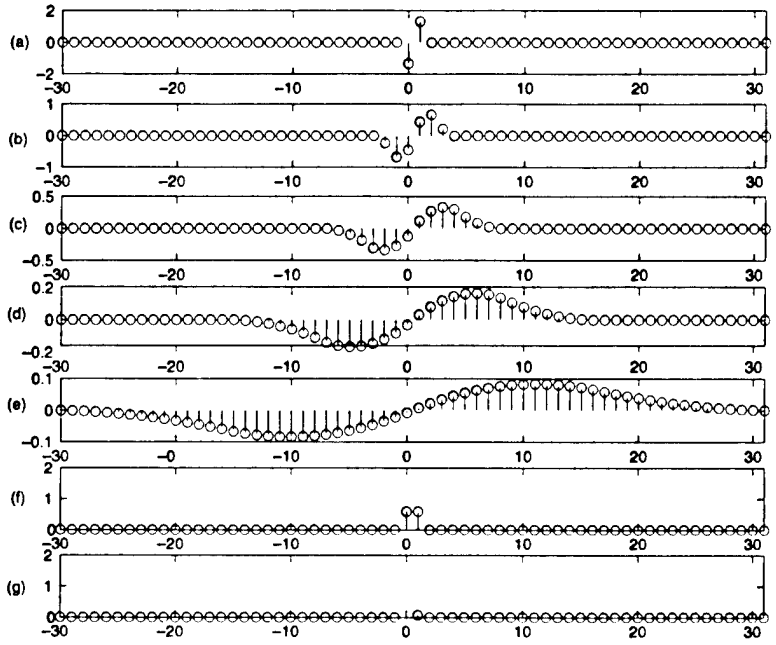
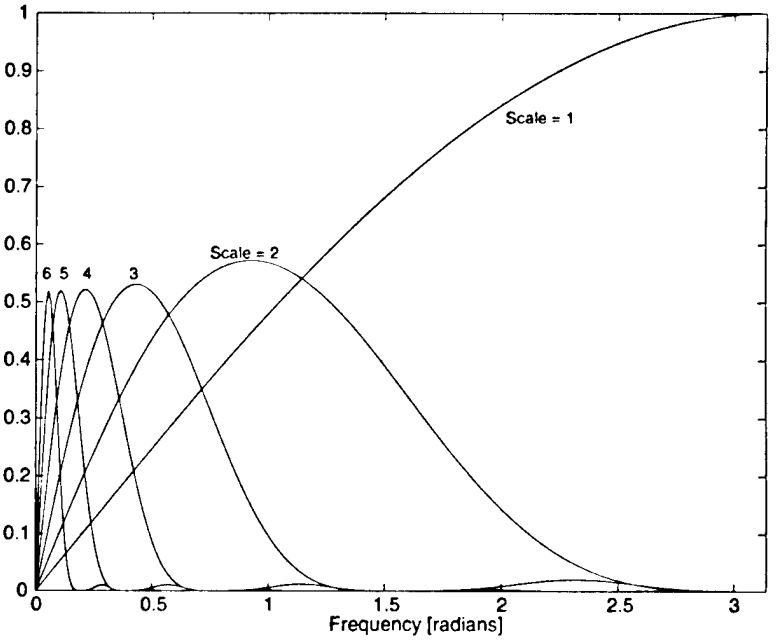


Figure 2. Frequency responses of DWT for first six scales.



by exploiting the multiscale correlation due to the presence of the desired signal. The multiscale product of the DWT outputs is given by

$$p(n) = \prod_{j=j_0}^{j_1} W_{2^j} x(n) = \prod_{j=j_0}^{j_1} y_j(n). \tag{9}$$

As further motivation, the number of local maxima due to noise decreases quickly with scale due to the increased smoothing. Specifically, for $x(n)$ white Gaussian noise the average number of local maxima at scale $s = 2^{j+1}$ is half that at scale $s = 2^j$ [21]. Thus, while maxima in $W_{2^j}x(n)$ due to edges in $x(n)$ will tend to propagate across scales, the maxima due to noise will tend not to, so that $p(n)$ will tend to reinforce the signal response and not the noise.

An example is shown in figure 3, depicting the noise-free step-response behavior of the MZ-DWT. Shown are (a) the time series, (b) through (e) the MZ-DWT for the first four scales, (f) $p(n)$ for $j_0 = 1$, and (g) $j_1 = 2$ and 3. The peak in the DWT at various scales corresponds to the step edge. Because the peaks align across the first few scales, the product $p(n)$ exhibits a corresponding peak.

A second example of the cross-scale product is given in figure 4. The signal is $x(n) = As(n) + v(n)$, with $A = 10$, $v(n)$ unit-variance white Gaussian noise, and $s(n)$ taking on values of 1 in the ranges $[51, 150]$, $[201, 250]$, $[301, 310]$, and $[361, 365]$, as well as impulses given by $\delta(n-416)$, $\delta(n-442)$, and $\delta(n - 445)$. The SNR is 20 dB, defined as $10 \log_{10}(A^2/\sigma_v^2)$. Here, peaks do not align across arbitrarily high scales because neighboring peaks interfere due to lengthening filter responses. The peaks in $p(n)$ are generally well pronounced, except for those corresponding to the isolated impulses between $n = 400$ and 450 in the original time series, where smoothing leads to weakened response at the higher scales (note also fig. 1 in this regard). This suppression of impulses and the effect of pulse width is explored in section 4. Peaks in figure 4(f) are generally positive going because of the even number of products, whereas those in figure 4(g) are bipolar and preserve the edge up/down direction information. In the following sections we develop properties and analyze behavior of $p(n)$.

Remarks

1. The inverse MZ-DWT may be readily accomplished [21,22]. In addition, reconstruction from local maxima via alternating projections may also be attained, leading to edge-detection-based filtering [8].

Figure 3. Noise-free step response of MZ-DWT: (a) time series, (b) through (e) first four scales of DWT, (f) $p_2(n)$, and (g) $p_3(n)$.

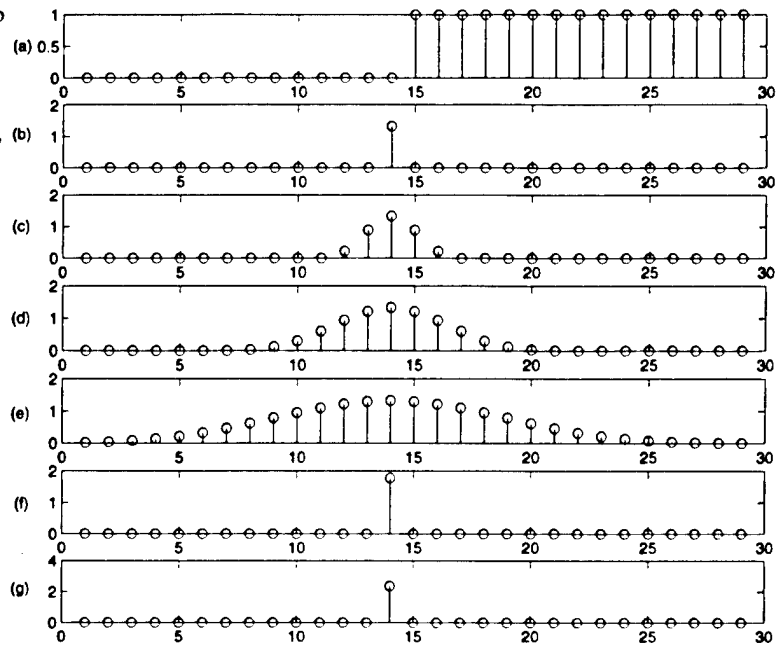
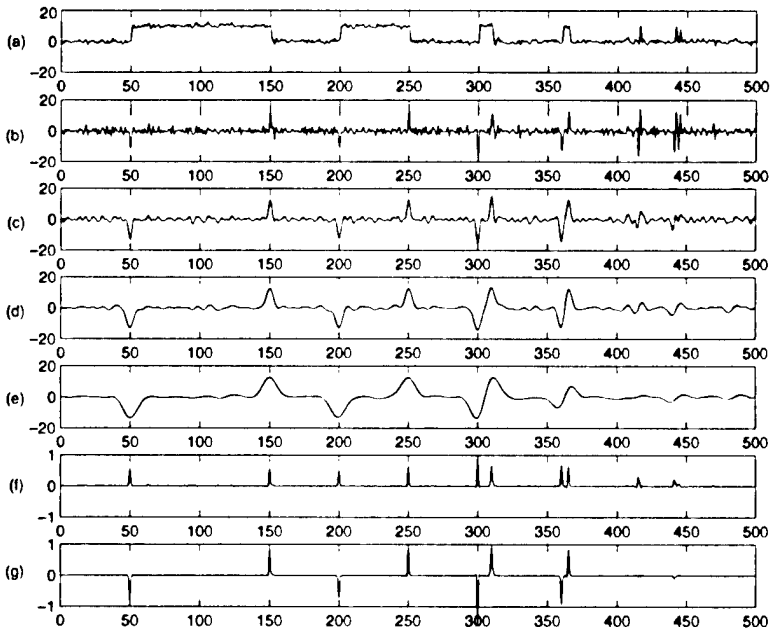


Figure 4. MZ-DWT example showing (a) time series, (b) through (e) first four scales of DWT, (f) normalized $p_2(n)$, and (g) normalized $p_3(n)$.



2. The region of support (RoS) of $h_{i+1}(n)$ is two plus twice the RoS of $h_i(n)$ and the RoS of $h_1(n) = 2$. These are tabulated in table 1; see also figure 1. Also, $h_i(n) = -h_i(-n)$, so that

$$\sum_{l_i} h_i(l_i) = 0, \quad \text{for all } i. \tag{10}$$

3. As previously noted, the problem can also be formulated in terms of zero-crossings of smoothed second derivatives (see, e.g., Mallat and Hwang [21]). Multiscale products could then be searched for minima.
4. To avoid discontinuities at the window edges when computing the MZ-DWT, we use an odd-symmetric extension of the data. The issue is discussed in Mallat and Zhong [22].
5. An interesting method for multiscale estimation is to use the iterative regularization approach [11,12], which might facilitate better alignment of multiscale peaks in the noisy signal case. This approach is nonlinear, and analysis along the lines presented here is consequently not straightforward.
6. Denjean and Castanie [9] have proposed multiscale detection tests. They form a multivariate Gaussian test statistic that uses various scales of a continuous wavelet transform.

Table 1. RoS of MZ-DWT FIR filters as function of scale.

Scale	RoS
1	2
2	6
3	14
4	30
5	62
6	126
7	254
8	510

3. A Nonlinear Whitening Transformation

In this section we consider the first- and second-order statistics (SOS) of the DWT. For numerical results we focus on the MZ-DWT implementation, while noting that the analysis applies more generally to filter banks with no downsampling, as previously described.

To begin our study of the DWT properties, consider the cross-correlation between outputs of the DWT. We assume that $x(n)$ is zero-mean real-valued white noise with correlation function $E[x(n)x(n+m)] = \sigma_x^2\delta(m)$. Using equation (4), we obtain

$$E[y_i(n)y_j(n)] = \sigma_x^2 \sum_k h_i(k)h_j(k). \tag{11}$$

It is straightforward to obtain the correlation coefficient between DWT outputs, given by

$$\rho_{ij} = \frac{\sum_k h_i(k)h_j(k)}{(\sum_{k_1} h_i^2(k_1) \sum_{k_2} h_j^2(k_2))^{1/2}}, \tag{12}$$

and the joint covariance matrix C of the DWT outputs at time n has (i,j) element given by

$$[C]_{ij} = \sigma_x^2 \sum_k h_i(k)h_j(k). \tag{13}$$

These relations are readily calculated for the MZ-DWT FIR filters. Correlation coefficients ρ_{ij} are shown at the top of table 2 for $1 \le i,j \le 8$, and the values on the diagonal of C are shown on the bottom. Note the decreasing correlation as the scales are separated, with (dyadic) scales a distance three apart being essentially uncorrelated. The uncorrelatedness of separated MZ-DWT outputs follows intuitively from the plot of MZ-DWT frequency responses shown in figure 2.

Next we consider properties of the multiscale product. Let

$$p_K(n) = y_1(n) \cdot y_2(n) \cdots y_K(n) \tag{14}$$

denote the Kth product of the outputs of the DWT, corresponding to $j_0 = 1$ and $j_1 = K$ in equation (9). For simplicity in the following we restrict ourselves to $j_0 = 1$, but the extension to the arbitrary case is straightforward. Assume now that $x(n)$ is zero mean and independent identically

Table 2. Top: Correlation coefficients ρ_{ij} between MZ-DWT outputs $y_i(n)$ and $y_j(n)$ (scales $s = 2^i$ and $s = 2^j$), for white noise input. Bottom: main diagonal values of covariance matrix C, with $\sigma_x^2 = 1$.

ρ_{ij}						
1.0000	0.5345	0.2097	0.0759	0.0270	0.0096	0.0034
0.5345	1.0000	0.6444	0.2791	0.1074	0.0395	0.0142
0.2097	0.6444	1.0000	0.6787	0.3019	0.1176	0.0435
0.0759	0.2791	0.6787	1.0000	0.6868	0.3075	0.1201
0.0270	0.1074	0.3019	0.6868	1.0000	0.6888	0.3089
0.0096	0.0395	0.1176	0.3075	0.6888	1.0000	0.6893
0.0034	0.0142	0.0435	0.1201	0.3089	0.6893	1.0000
$[C]_{ii}$						
3.5556	1.1905	0.3236	0.0825	0.0208	0.0052	0.0013

distributed (iid). The mean is given by

$$\begin{aligned} E[p_K(n)] &= E[y_1(n)\cdots y_K(n)] \\ &= \sum_{l_1} h_1(l_1)\cdots\sum_{l_K} h_K(l_K)E[x(n-l_1)\cdots x(n-l_K)]. \end{aligned} \tag{15}$$

It follows that $E[p_1(n)] = 0$, and $E[p_3(n)] = 0$ due to the anti-symmetry of $h_j(n)$ (see eq (10)); the latter also follows if $x(n)$ is symmetrically distributed such as Gaussian. More generally, $E[p_K(n)] = 0$ for K odd if the odd order cumulants of $x(n)$ are zero.

In general, $E[p_K(n)] \neq 0$ for K even. For $K = 2$, invoking the whiteness assumption on $x(n)$ yields, using table 2,

$$E[p_2(n)] = \sigma_x^2 \sum_i h_1(i)h_2(i) = 1.1905\sigma_x^2. \tag{16}$$

For $K = 4$, expanding the fourth-moment in equation (15) yields

$$E[p_4(n)] = \gamma_{4x}\ \sigma_x^4 \left[\sum_l h_1(l)h_2(l)h_3(l)h_4(l)] + \sigma_x^4\ [r_{12}(0)r_{34}(0) + r_{13}(0)r_{24}(0) + r_{14}(0)r_{23}(0)\right], \tag{17}$$

where

$$r_{ij}(\tau) = \sum_l h_i(l)h_j(l+\tau) \tag{18}$$

is the deterministic correlation of the filter IRs and

$$\gamma_{4x} = \frac{E[x^4(n)]}{E[x^2(n)]^2} - 3 \tag{19}$$

is the normalized kurtosis. Evaluation of equation (17) for the MZ-DWT yields

$$E[p_4(n)] = 0.0045\ \gamma_{4x}\ \sigma_x^4 + 0.4940\ \sigma_x^4 = 0.0045\ \sigma_x^4\ [\gamma_{4x} + 110.50]. \tag{20}$$

Note from equation (20) that unless γ_{4x} is large, the SOS-related terms dominate the mean for $K = 4$. If $x(n)$ is Gaussian, then $\gamma_{4x} = 0$ and equation (20) simplifies accordingly.

We find the autocorrelation of $p_K(n)$ as follows. We have

$$\begin{aligned} r_{p_K}(\tau) &= E[p_K(n)p_K(n+\tau)] \\ &= E\left[\prod_{i=1}^{K}\left(\sum_{l_i} h_i(l_i)x(n-l_i)\right)\prod_{j=1}^{K}\left(\sum_{m_j} h_j(m_j)x(n+\tau-m_j)\right)\right] \\ &= \sum_{l_1} h_1(l_1)\cdots\sum_{l_K} h_K(l_K)\sum_{m_1} h_1(m_1)\cdots\sum_{m_K} h_K(m_K) \\ &\quad \times E\left[\prod_{i=1}^{K} x(n-l_i)\prod_{j=1}^{K} x(n+\tau-m_j)\right]. \end{aligned} \tag{21}$$

To evaluate equation (21), let us now assume that $x(n)$ is Gaussian, so that the expectation in equation (21) is a $(2K)$th joint moment of Gaussians. This even-order moment reduces to a sum of $1 \times 3 \times \cdots \times (2K-1)$ terms, where each term is composed of appropriate products of $r_{ij}(0)$ and $r_{ij}(\tau)$, defined via equation (18). For $K = 2$ the expectation in equation (21) reduces to three terms, and using the whiteness of $x(n)$ we have

$$r_{p_2}(\tau) = \sigma_x^4[r_{12}^2(0) + r_{11}(\tau)r_{22}(\tau) + r_{21}(\tau)r_{12}(\tau)]. \tag{22}$$

Similarly, for $K = 3$, the expectation reduces to a sum of 15 product terms in $r_{ij}(\tau)$, given by

$$\begin{aligned} r_{p_3}(\tau) &= \sigma_x^6[r_{12}(0)\cdot r_{31}(\tau)\cdot r_{23}(0) + r_{12}(0)\cdot r_{32}(\tau)\cdot r_{13}(0) + r_{12}(0)\cdot r_{33}(\tau)\cdot r_{12}(0) \\ &\quad + r_{13}(0)\cdot r_{21}(\tau)\cdot r_{23}(0) + r_{13}(0)\cdot r_{22}(\tau)\cdot r_{13}(0) + r_{13}(0)\cdot r_{23}(\tau)\cdot r_{12}(0) \\ &\quad + r_{11}(\tau)\cdot r_{23}(0)\cdot r_{23}(0) + r_{11}(\tau)\cdot r_{22}(\tau)\cdot r_{33}(\tau) + r_{11}(\tau)\cdot r_{23}(\tau)\cdot r_{32}(\tau) \\ &\quad + r_{12}(\tau)\cdot r_{23}(0)\cdot r_{13}(0) + r_{12}(\tau)\cdot r_{21}(\tau)\cdot r_{33}(\tau) + r_{12}(\tau)\cdot r_{23}(\tau)\cdot r_{31}(\tau) \\ &\quad + r_{13}(\tau)\cdot r_{23}(0)\cdot r_{12}(0) + r_{13}(\tau)\cdot r_{21}(\tau)\cdot r_{32}(\tau) + r_{13}(\tau)\cdot r_{22}(\tau)\cdot r_{31}(\tau)]. \end{aligned} \tag{23}$$

The number of terms grows quickly, making the algebra somewhat tedious; for example, $K = 4$ results in 105 terms.

For the MZ-DWT, $r_{p_K}(m)$ may be easily calculated for the various cases. Plots of $r_{p_2}(\tau)$ and $r_{p_3}(\tau)$ are shown in figure 5. Note the dominance of $r_{p_2}(0)$ and $r_{p_3}(0)$. For $K \geq 3$, $r_{p_K}(\tau)/r_{p_K}(0) \approx \delta(\tau)$. This effect becomes more pronounced as K is increased, as confirmed by analysis and simulations. Despite both the temporal dependence in each of the DWT time

series $y_k(n)$, as well as the cross-correlation between these outputs, $p_K(n)$ is effectively whitened for $K \geq 3$. This is intuitively apparent from figure 2, because the time domain product corresponds to convolution in the frequency domain.

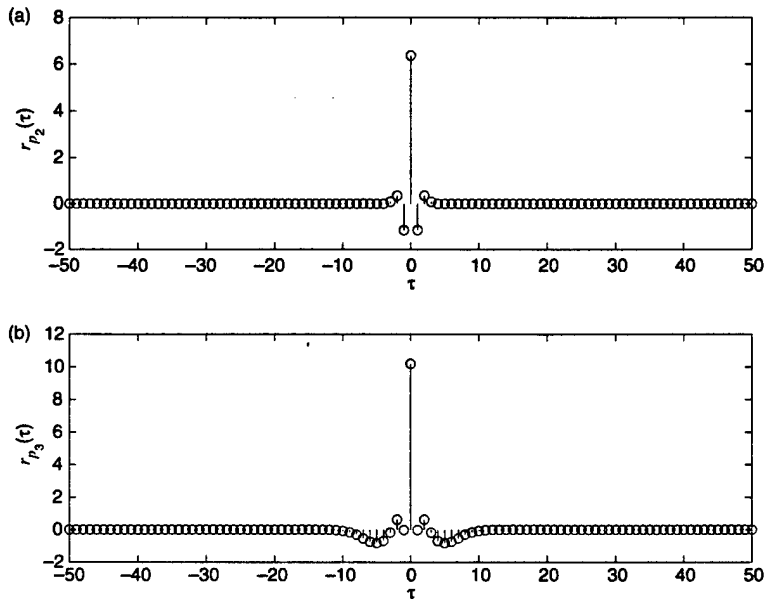


Figure 5. Theoretical correlations of (a) $p_2(n)$ (with bias removed) and (b) $p_3(n)$, depicting whitening effect of multiscale product.

Remarks

1. In deriving $r_{p_K}(\tau)$, we can handle the case of $x(n)$ non-Gaussian in a manner similar to that above by expressing the moment in equation (21) in terms of cumulants and exploiting the assumed iid nature of $x(t)$. This will also generally involve higher-order analogs of equation (18). For example, for $K = 2$ with $x(n)$ iid non-Gaussian, we obtain

$$r_{p_2}(\tau) = \sigma_x^4[r_{12}^2(0) + r_{11}(\tau)r_{22}(\tau) + r_{21}(\tau)r_{12}(\tau)] + \gamma_{4x}\,\sigma_x^4 \left[\sum_l h_1(l)h_2(l)h_1(l+\tau)h_2(l+\tau)\right], \quad (24)$$

which corresponds to the Gaussian case solution of equation (22) plus an additional term that is proportional to γ_{4x}. Evaluating for the MZ-DWT, we find that the second term in equation (24) is $\gamma_{4x}\,\sigma_x^4[0.7086\delta(\tau) + 0.3543\delta(\tau \pm 1)]$.

4. Step and Pulse Response

Having shown that forming multiscale products yields a whitening transformation with respect to noise, we next consider the response to deterministic steps and pulses of varying widths. This enables us to characterize the SNR in the multiscale product domain, and to examine the effects of smoothing on multiscale-product edge response as a function of the pulsewidth. The latter is particularly important to understand before applying the multiscale product approach.

It is natural to define an SNR in the multiscale product domain. Suppose the signal is given by $s(n) = Au(n-1) + x(n)$, where $x(n)$ is additive white Gaussian noise as above, $A > 0$ a positive constant, and $u(n)$ is the unit step function; $u(n) = 0$, $n < 0$, and $u(n) = 1$, $n \geq 0$. We may define the input SNR in decibels as

$$\mathrm{SNR}_i = 10 \log_{10} \frac{A^2}{\sigma_x^2}. \tag{25}$$

The (noise-free) multiscale product response to $u(n-1)$ at the step-change time $n = 1$ is

$$p_K(n) = \prod_{i=1}^{K} y_i(n) = A^K \prod_{i=1}^{K} \left(\sum_i h_i(l_i) u(n - 1 - l_i) \right). \tag{26}$$

In this case the MZ-DWT yields $p_K(0) \approx (1.33A)^K$, for all scales (the MZ-DWT filters have been designed so as to yield the same noise-free step response amplitude at all scales; see fig. 3). We can now define the output SNR as

$$\mathrm{SNR}_o = 10 \log_{10} \frac{[p_K(0)|_{u(n-1)}]^2}{r_{p_K}(0)|_{x(n)}}, \tag{27}$$

where $p_K(0)|_{u(n-1)}$ denotes the peak step response and $r_{p_K}(0)|_{x(n)}$ is the variance of $p_K(n)$ due to $x(n)$. Note that

$$\frac{[p_K(0)|_{u(n-1)}]^2}{r_{p_K}(0)|_{x(n)}} = \beta \left(\frac{A^2}{\sigma_x^2} \right)^K; \tag{28}$$

so $\mathrm{SNR}_o \propto (\mathrm{SNR}_i)^K$ and the Kth-order product nonlinearity results in amplification of SNR_i. The SNR input-output relationship is plotted in

figure 6, for $K = 2$ and $K = 3$, using equations (22) and (26). The line $\text{SNR}_i = \text{SNR}_o$ is shown for reference. There is SNR gain for SNR_i sufficiently positive, whereas there is an SNR loss for lower values of SNR_i; and the SNR gain increases as SNR_i increases. Here $\beta = 0.4014$ for $K = 2$, and $\beta = 0.5428$ for $K = 3$, so the loss-gain thresholds are 4 dB ($K = 2$) and 1.33 dB ($K = 3$).

The above describes the response to an (essentially infinite) step function. Next we consider the response to a pulse of finite width. In this case the interaction between edges becomes an issue when they are both contained within the impulse response region of support. Also, the smoothing at higher scales will smear short pulses and dampen the maxima in $p_K(n)$, as depicted in figure 4. This is evident from study of the MZ-DWT IRs in figure 1, whose values in the neighborhood of occurrence of the impulse grow steadily smaller with increasing scale. The suppression of small features in $p_K(n)$ is a direct result of the smoothing at the higher scales, and may be considered beneficial or harmful, depending on one's point of view. These effects can be quantified by examining the noise-free response to a pulse of varying width. Figure 7(a) shows normalized values of $p_3(0)$, where the input is a pulsewidth of n_0, given by $s(n) = u(n-1) - u(n-1-n_0)$, for $n_0 = 1, \cdots, 11$. Here $p_3(0)$ is the peak response to the leading edge of the pulse. We plot $10 \log_{10} p_3(0)$, normalized by the maximum of $p_3(0)$ over the range of n_0. Pulsewidths of $n_0 \geq 6$ yield essentially the same peak response in $p_3(n)$, equivalent to the ideal unit-step response. The case of $n_0 = 1$ corresponds to $s(n) = \delta(n-1)$, an impulse. Figure 7 indicates a (deterministic)

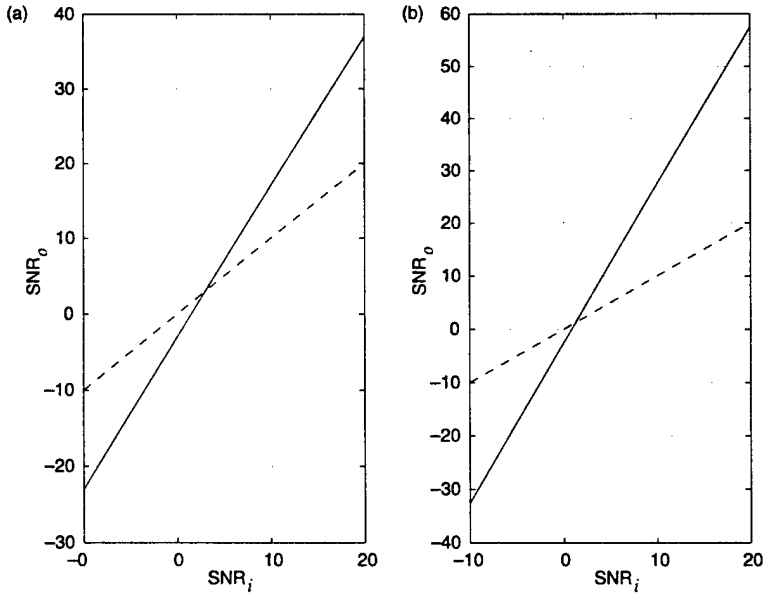


Figure 6. SNR input-output relationship for step input in white Gaussian noise: (a) for $p_2(0)$, and (b) for $p_3(0)$. Dashed line $\text{SNR}_i = \text{SNR}_o$ is shown for reference.

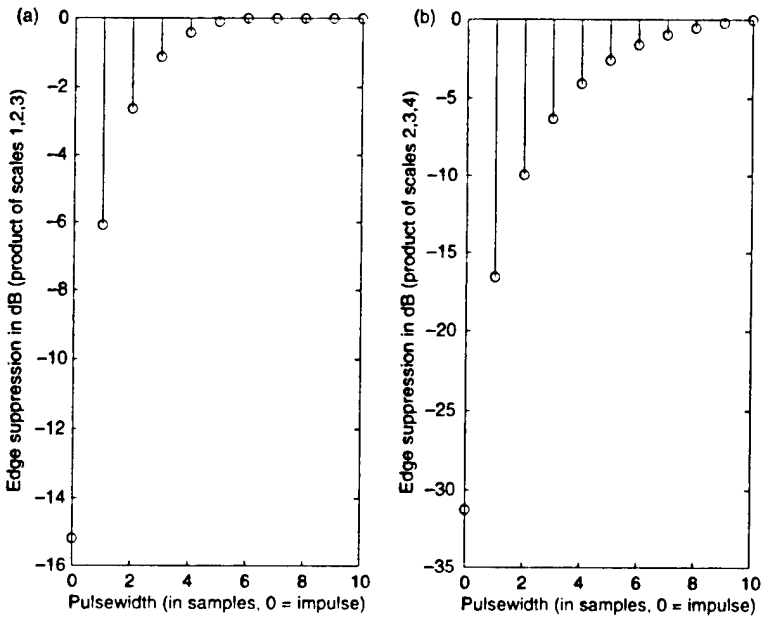


Figure 7. Edge suppression: (a) product of scales 1, 2, and 3, and (b) product of scales 2, 3, and 4.

15-dB suppression of impulses in $p_3(n)$, relative to the noise-free step response that is achieved for a pulsewidth of $n_0 \geq 6$. The figure also shows $\approx$ 3 dB suppression of edge response for a pulsewidth of 2 (three samples total).

We repeat the above, now using the product of scales 2, 3, and 4, with results shown in figure 7(b). In this case, the additional smoothing and omission of the first scale leads to more suppression of impulses relative to the step response (about 31 dB). However, the price paid is that a pulsewidth of 10 or more samples is now required for the edge response to match that of the ideal step response, and pulsewidths of 5 have $\approx$ 3 dB suppression in edge response.

While the results of this and the previous section are encouraging, they do not tell the entire story. In the following section, we examine the pdf of multiscale products, and find them to be strongly heavy-tailed non-Gaussian. This has implications for edge detection in the multiscale product domain.

5. Multiscale Products are Heavy Tailed

Next we study the univariate pdf's of the multiscale products $p_2(n)$ and $p_3(n)$. We show that they are in general non-Gaussian and heavy tailed. We then consider detection, and use the pdf's to set detection thresholds.

5.1 pdf's of Multiscale Products

Although it is whitened with respect to its second-order statistics, $p_K(n)$ is in general distinctly non-Gaussian. However, determination of the first-order pdf is not straightforward, even with $x(n)$ Gaussian. A closed form is only available for the bivariate case, given by Miller [24, sect. 2.3]:

$$f(z) = \frac{1}{\pi}\sqrt{|W|}\, e^{-w_{12}z}\, K_0\left(|z|\sqrt{w_{11}w_{22}}\right), \tag{29}$$

where $W = C^{-1}$ is the inverse of the bivariate covariance matrix with elements $[w]_{ij}$, $1 \le i,j \le 2$, and K_0 is the modified Bessel function of the second kind and order zero. For the MZ-DWT we find that for the first two scales

$$W = C^{-1} = \begin{bmatrix} 3.5556 & 1.1905 \\ 1.1905 & 1.3951 \end{bmatrix}^{-1} = \begin{bmatrix} 0.3937 & -0.3360 \\ -0.3360 & 1.0035 \end{bmatrix}, \tag{30}$$

and $|W| = 0.2822$; so that in this case

$$f(z) = \frac{0.5313}{\pi}\, e^{0.3360z}\, K_0(0.6286\,|z|). \tag{31}$$

For third or higher order products, we must resort to parametric or non-parametric numerical techniques, as we describe next.

Figure 8 depicts pdf's for two cases with $x(n)$ white Gaussian, the theoretical pdf for $K = 2$ from equation (31), and an unsmoothed histogram estimate for $K = 3$ based on 5×10^6 samples. Unit-variance Cauchy and normal pdf's are shown for reference. Note the skewness in the $K = 2$ case, due to the positive correlation between $y_1(n)$ and $y_2(n)$, and because K is even. In this case the positive tail is heavy, showing a strong impulsive nature in $p_2(n)$ on the positive side, with a much lighter tail on the negative side. Experiments with numerical estimation of the pdf in the $K = 2$ case exhibit

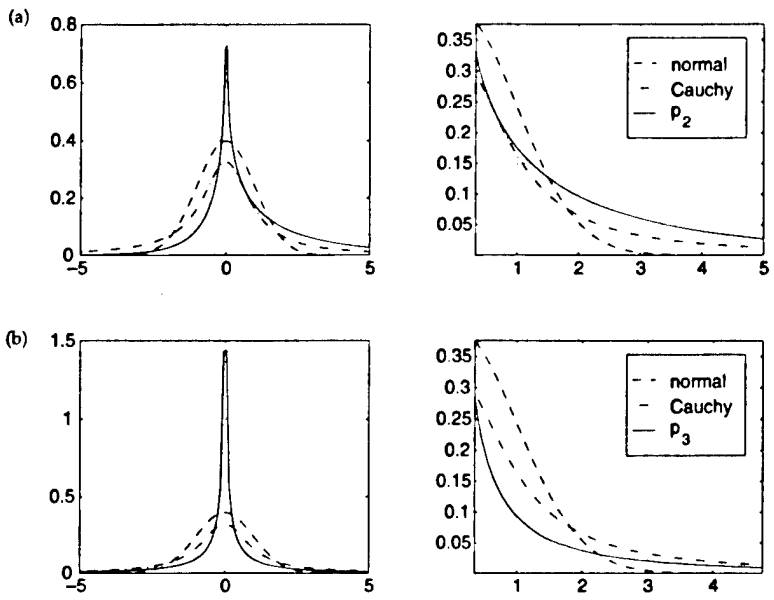


Figure 8. (a) Theoretical first-order pdf of $p_2(n)$ and (b) estimated first-order pdf of $p_3(n)$. Unit variance normal and Cauchy shown for reference.

excellent agreement with the theory. In the $K = 3$ case the pdf is symmetric with tail behavior heavier than Gaussian but lighter than Cauchy.

An appealing alternative to histogram or kernel-based pdf estimation is to employ an L-term Gaussian mixture model, given by

$$f(z) = \sum_{l=1}^{L} \frac{\lambda_l}{2\pi\sigma_l^2} \exp\left(-\frac{|z|^2}{2\sigma_l^2}\right), \tag{32}$$

where the model is parameterized by $\Sigma = [\sigma_1^2, \cdots, \sigma_L^2]$ and $\Lambda = [\lambda_1, \cdots, \lambda_{L-1}]$, with $\lambda_L = 1 - \sum_{l=1}^{L-1} \lambda_l$. The model encompasses a broad range of symmetric zero-mean pdf's; e.g., see Liporace [19]. Approximate maximum-likelihood (ML) estimates of the parameters are readily obtained using the iterative expectation-maximization (EM) algorithm; e.g., see McLachlan and Krishnan [23, sect. 2.7]. As is well known, each iteration of the EM algorithm achieves nondecreasing likelihood, although local convergence may be the result, depending on initialization. However, for heavy-tailed symmetric pdfs (all mixture terms zero mean, as in eq (32)) very good results can be obtained for small sample sizes with minimal dependence on initialization. The Gaussian mixture is readily applied to detection and estimation problems in heavy-tailed noise [5,16,17].

Figure 9 shows an example of $\hat{f}(z)$ for $p_3(n)$, obtained from 1000 samples and $L = 10$ in equation (32). The EM algorithm was iterated 100 times; this was more than adequate in our tests. For initialization we used

$$\lambda_1 = 0.98, \lambda_2 = 0.005, \lambda_i = 0.0025 \text{ for } 3 \leq i \leq 6, \lambda_i = 0.00125 \text{ for } 7 \leq i \leq 10,$$
$$\Sigma = \hat{\sigma}_o^2[1, 5, 10, 15, 20, 100, 500, 1000, 2000, 4000],$$

where $\hat{\sigma}_o^2$ is a robust estimate of the variance given by [20, 400]

$$\hat{\sigma}_o^2 = \left(\frac{\overline{\sigma}}{0.7}\right)^2, \tag{33}$$

where $\overline{\sigma}$ (the median absolute deviation) is the median of $|x(n) - \tilde{x}|$, and $\tilde{x}$ is the median of the observations $x(n)$. The final parameter estimates of $\hat{f}(z)$ in equation (32) are shown in table 3. The overshoot near the origin and undershoot in the estimated pdf tail are typical of our experience, apparently due to the sample size. We have obtained useful estimates with as few as 100 samples because of the relatively heavy tails, i.e., a sample size of 100 yields enough information about tail behavior to form meaningful pdf estimates.

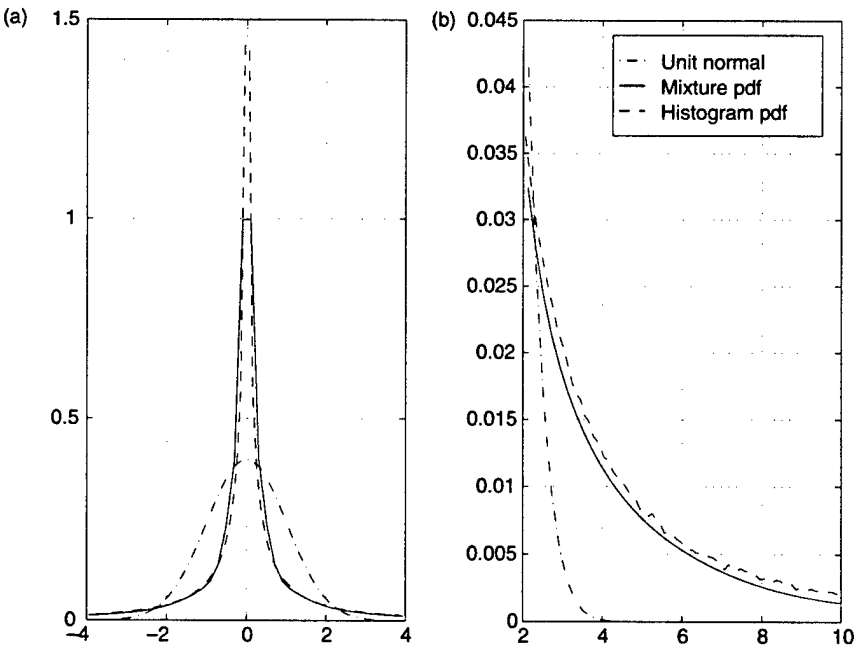


Figure 9. Estimated first-order pdfs of $p_3(n)$ comparing nonparametric large sample histogram with parametric ten-term Gaussian mixture. (Panel (b) is a blowup of panel (a).) Mixture parameters estimated via EM algorithm: 1000 samples and 100 iterations.

In table 4 we show some estimated statistics of $p_K(n)$ for $1 \leq K \leq 4$, for iid unit-variance Gaussian input and iid unit-variance Laplace input, including normalized skewness and normalized kurtosis. Assuming zero-mean, normalized kurtosis is defined in equation (19), and the normalized skewness is given by

$$\gamma_{3x} = \frac{E[x^3(n)]}{E[x^2(n)]^{3/2}} \,. \tag{34}$$

Table 3. Estimated parameters of ten-term Gaussian mixture approximation to non-Gaussian pdf of $p_3(n)$ with white Gaussian input to MZ-DWT.

l	λ_l	σ_l^2
1	0.0709	0.0000
2	0.2429	0.0035
3	0.2914	0.1031
4	0.1562	0.8872
5	0.0940	3.0992
6	0.1107	14.0930
7	0.0136	121.0198
8	0.0102	128.3395
9	0.0064	128.6320
10	0.0036	128.6911

Table 4. Estimated parameters of $p_K(n)$ for varying K, showing heavy-tailed nature of pdf's, and skewness for K even.

Unit-variance Gaussian input				
K = No. of products	1	2	3	4
Mean	0.0000	1.1899	0.0004	0.4948
Variance	3.5531	6.3654	10.1739	6.9256
Skewness	0.0007	2.4063	0.0143	10.4348
Kurtosis	–0.0010	10.1278	52.7261	230.4220
Unit-variance Laplace input				
K = No. of products	1	2	3	4
Mean	0.0000	2.3805	0.0029	2.0214
Variance	2.6666	5.8274	10.9141	13.3247
Skewness	–0.0003	4.0264	–0.1553	18.0487
Kurtosis	1.4981	32.0013	171.7648	825.0208

These estimates were obtained by averaging over 100 runs, with 10^5 samples for each run. They demonstrate the heavy-tailed nature of $p_K(n)$, as well as the strong skewness for K even. These may also be evaluated theoretically, although the task is tedious as K increases.

5.2 Detection in the Multiscale Product Domain

Knowledge of the pdf of $p_K(n)$ may be used to set detection thresholds to achieve a constant false alarm rate. This can be achieved with a known or estimated noise variance. We show a 1-D edge detection example in figure 10, where detection is declared when $|p_3(n)|$ exceeds a threshold. The threshold was obtained from the histogram estimate of the pdf for $p_3(n)$, and set to achieve $P_{fa} = 0.01$ (the resulting threshold was 13.5053). Three different edges were used in additive white Gaussian noise with variance $\sigma_v^2 = 1$. Consider a sigmoidal step change signal, given by

$$x_s(n) = \frac{m_2 + m_1 e^{-\alpha T(n-n_0-\tau)}}{1 + e^{-\alpha T(n-n_0-\tau)}} \,. \tag{35}$$

This model was used by Reza and Doroodchi [25]. The parameters m_1 and m_2 are the signal levels before and after the step, T is the sampling interval time, parameter α determines the risetime, and the step occurs (in continuous time) at $n_0 T + \tau$. Without loss of generality we can assume $T = 1$, because T can always be incorporated into α. Let $A = m_2 - m_1$ denote the step change. An alternative expression for $x_s(n)$ in terms of the tanh function is

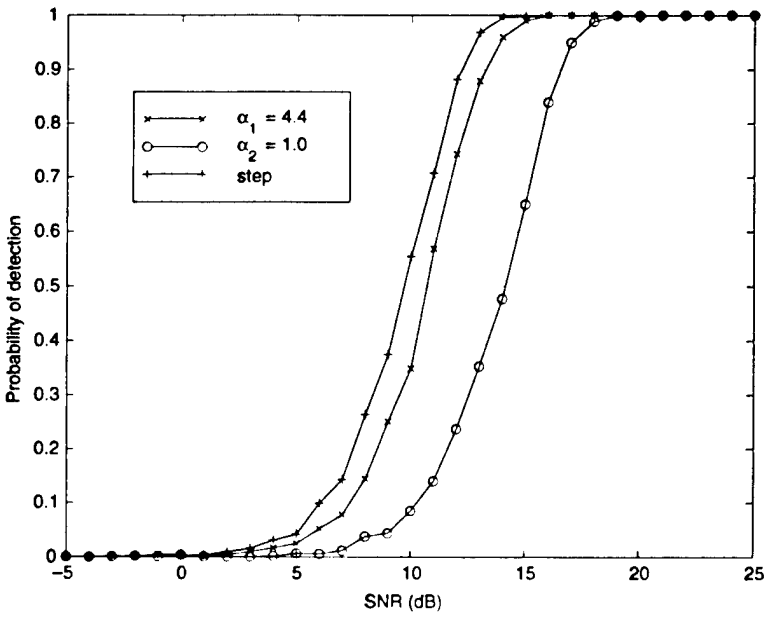
$$x_s(n) = \frac{A}{2}[1 + \tanh\left(\frac{\alpha T}{2}(n - n_o - \tau)\right)] + m_1, \quad n = 0, ..., N-1 \,. \tag{36}$$

For the example of figure 10, two of the signals were based on $x_s(n)$. The parameter τ was uniformly distributed in $[0, T]$, modeling the effects of time quantization via sampling. The SNR is defined in decibels as

$$\text{SNR} = 10 \log_{10} \frac{A^2}{\sigma_v^2} \,. \tag{37}$$

Two cases of equation (36) were implemented, $\alpha = 4.4$ and $\alpha = 1$, corresponding to a fast and slower step risetime, respectively. In both cases $m_1 = 0$ and $T = 1$, and m_2 was adjusted to achieve the desired SNR. The third curve in figure 10 corresponds to a unit step, from $m_1 = 0$ to m_2, where m_2 is set to achieve the desired SNR. One thousand Monte Carlo trials were run for each SNR value, and $p_3(n)$ was tested at the position where the edge should produce a maximum. As we would expect, slower risetime results in poorer detection performance. We also note the moderate to high SNR required for a high probability of detection.

Figure 10. Probability of step detection using $p_3(n)$ in white Gaussian noise for three different edge shapes with $P_{fa} = 0.01$.



6. Estimation of Step-Change Location

Next we consider estimation of step-change location. This assumes that a change has occurred in the observation interval, hence there is no detection problem. In the appendix we develop a general-closed form CRB for step-change estimation in additive iid non-Gaussian noise, and provide specific expressions for the step-change signal model given by equation (35) or (36). This generalizes the discrete-time results of Reza and Doroodchi [25].

The noisy discrete-time signal is given by

$$y(n) = x(n) + v(n) = s(\alpha(n - n_o - \tau)) + v(n), \quad n = 0, ..., N-1 \qquad (38)$$

where $v(n)$ is zero-mean white Gaussian noise with variance σ_v^2. As described in the previous section, the step-change is centered at time $n_o + \tau$, where τ is a sample-time , and α is a scaling parameter that determines the risetime of the step (larger α corresponds to a faster risetime). A general expression for the Fisher information is given in the appendix by equation (A-8) for τ random and uniformly distributed over the sampling interval; an asymptotic expression is also given in equation (A-9). These expressions indicate that the Fisher information is proportional to the risetime parameter α, and inversely proportional to the noise variance.

To quantify the general case, we consider the step change signal model of equation (35) or (36). For this sigmoidal step model we find the closed-form Fisher information to be equation (A-13); an asymptotic form is also provided in equation (A-14). Figure 11 plots the Fisher information of equation (A-13) as a function of αTN and d, where T is the sampling time, N is the window size, and $d = n_o/N$ is the fractional index of the location parameter (see the discussion in the appendix). Thus $d \approx 0.5$ corresponds to the step change occurring approximately at the center of the observation window, and larger values of αTN indicate that the signal variation is more completely captured within the observation window. From figure 11 we see that the Fisher information reaches its asymptotic maximum for a range of d of approximately $0.2 \leq d \leq 0.8$, for $\alpha TN > 15$. Even for $\alpha TN = 2$ the Fisher information is about half its asymptotic value. This indicates that only local information around the step change is necessary to achieve good estimation of location.

It is also interesting to consider the case when τ is a nonrandom fixed constant τ_0. The Fisher information for this case is given in the appendix in

equation (A-11). We plot this case in figure 12, showing Fisher information versus parameter αT, with curves parameterized by $0 \leq \tau \leq 0.5$. Due to symmetry the case of $\tau = 0.6$ is equivalent to that for $\tau = 0.4$, and so on. The case of $\tau = 0$ (or $\tau = 1$) corresponds to the step-change being centered exactly at the sample time, so that the Fisher information increases without bound as αT increases. The curves for $\tau \neq 0$ in figure 12 show the sensitivity to τ in estimating step location. Depending on the step risetime

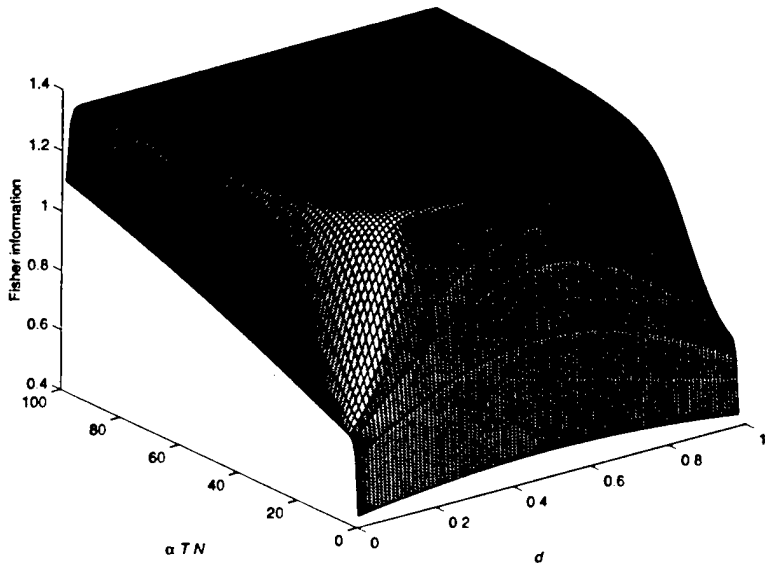


Figure 11. Fisher information for step-change location estimation with uniformly distributed sample-time offset τ.

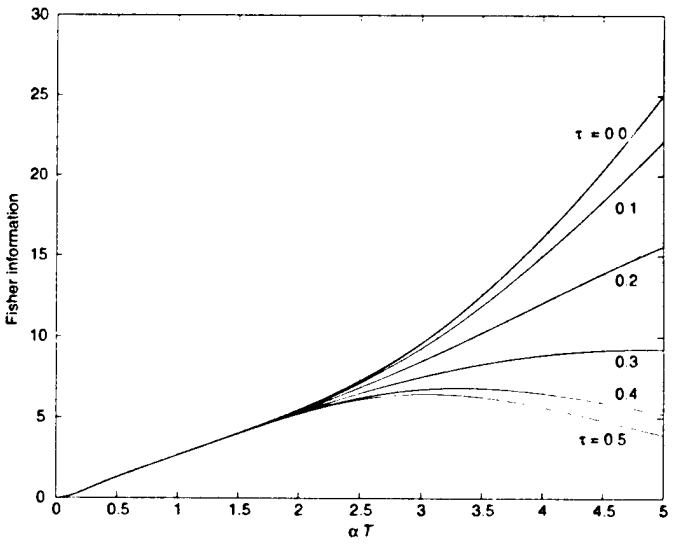


Figure 12. Fisher information for step-change location estimation with sample-time offset τ nonrandom. Curves parameterized by τ; $\tau = 0$ implies sample time and step-change aligned, $\tau = 0.5$ implies step-change center exactly between samples.

and the sampling time, it matters where the samples fall on the continuous representation of the step. We note that as $\alpha T \to \infty$, the step risetime becomes instantaneous; in this limit all of the curves decrease to a limiting value (except the $\tau = 0$ and $\tau = 1$ cases) as the samples will bracket the instantaneous change and no sub-sample time accuracy can be determined.

As an example, we show in figure 13 the theoretical CRB, as well as experimental mean-square error (MSE), for two edge-location estimation methods. The step was modeled via equation (35) with τ uniformly distributed. The two estimation methods are first, based on $p_3(n)$, and second, based on a simple gradient estimator with FIR given by $[-1, 0, 1]$. The sigmoidal function was generated with $m_1 = 0$, $T = 1$, and $\alpha = 4.4$, corresponding to a risetime of about $4T$ (see Reza and Doroodchi [25]). The step height m_2 was set to achieve the desired SNR. We used $N = 256$ and averaged over 5000 Monte Carlo trials for each SNR value. For each trial τ was selected uniformly in $[-0.5, 0.5]$. The step was centered at $N/2$, and the estimate of step location was taken to be the maximum in a 32-sample observation window centered on $N/2$. The simple gradient estimator has no smoothing, whereas $p_3(n)$ exploits multiple smoothing levels. At low SNR the benefit of the smoothing is evident, as the DWT method outperforms the simple gradient estimator by about 5 dB. At high SNR both methods are time-quantization error dominated. The simple gradient approaches the minimum variance of $1/12$ of a sampling time, which arises due to the uniform distribution of the sampling phase error. The DWT method approaches $1/2$, reflecting the effects of smoothing via the increased variance of the estimate.

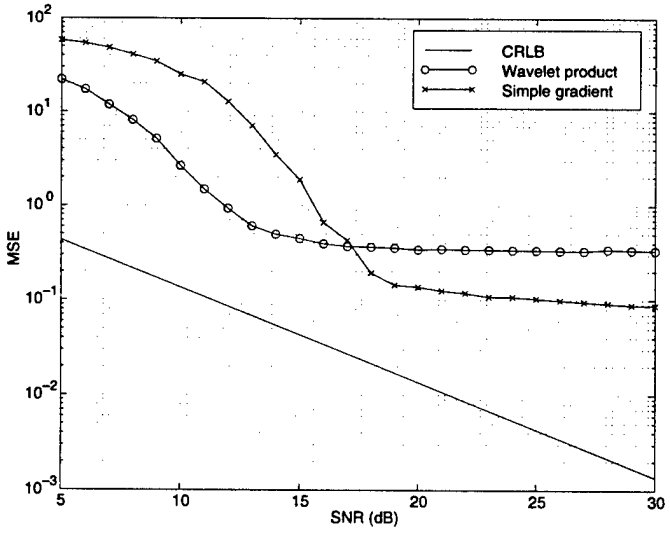


Figure 13. MSEs and theoretical CRB for step-change location estimation.

Changing α changes the effective risetime of the sigmoidal signal. Decreasing α creates a slower rising step. This results in a shift upward of the CRB as the step location is inherently more difficult to estimate. Repeating the experiment in figure 13 with lower α results in a shift to the right as higher SNR is needed to obtain the same performance, while the general behavior of the curves is the same.

Remarks

1. In forming multiscale products, the individual outputs $y_i(n)$ could be raised to different powers before the product operation. Alternatively, one could search for minima in the reciprocal multiscale product. An interesting special case arises in products of the form

$$X_m = \prod_{j=1}^{m} (Y_j)^{-2j}. \tag{39}$$

   For Y_j iid Gaussian, X_m is α-stable, with $\alpha = 2^{-m}$, and skewness parameter $\beta = 1$ [6]. With these values for α, X_m has finite moments only for orders less than 2^{-m}, which decreases rapidly as m increases, resulting in very heavy-tailed distributions for X_m. The case of $m = 1$ is well known, corresponding to the Pearson V or Levy pdf. One could also use $\log X_m$. Applying this result to MZ-DWT outputs would require skipping scales in order to approximate the iid assumption (see table 2).

7. Discussion

We have analyzed the use of products for nonlinearly combining multiscale wavelet outputs for detecting and estimating steps and edges. The smoothed gradient DWT developed by Mallat and Zhong provides a wavelet framework for this approach. The technique exploits the low complexity of the DWT computation, amounting to a few FIR filters. Despite its nonlinear nature the multiscale product is whitening, although the resulting noise pdf is distinctly non-Gaussian and generally heavy tailed. Closed-form expressions are unavailable for products of order greater than two, but pdf estimates may be obtained in a relatively easy manner. The heavy-tailed nature of the multiscale products is problematic from a detection viewpoint, as edge detection in the multiscale-product domain amounts to detecting impulses in impulsive noise. This points to the more general idea that gradient estimation is not particularly well posed, and that these methods are likely to be used for detection only in a moderate to high SNR regime as, for example, in image processing problems. Location estimation may be characterized by comparison with the general CRB. The results indicate tradeoffs possible in exploiting multiple smoothing levels when the most appropriate single smoothing level is not known a priori.

References

1. M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes, Theory and Application* (Prentice-Hall, 1993).

2. M. Basseville, B. Espiau, and J. Gasnier, "Edge detection using sequential methods for change in level—Part I: A sequential edge detection algorithm," *IEEE Trans. Acoust., Speech, and Sig. Proc.*, Vol. ASSP-29, No. 1, pp. 24–31, 1981.

3. M. Basseville, "Edge detection using sequential methods for change in level—Part II: Sequential detection of change in mean," *IEEE Trans. Acoust., Speech, and Sig. Proc.*, Vol. ASSP-29, No. 1, pp. 32–50, 1981.

4. M. Basseville and A. Benveniste, "Design and comparative study of some sequential jump detection algorithms for digital signals," *IEEE Trans. Acoust., Speech, and Sig. Proc.*, Vol. ASSP-31, No. 3, pp. 521–535, 1983.

5. R. S. Blum, R. J. Kozick, and B. M. Sadler, "An adaptive spatial diversity receiver for non-Gaussian interference and noise," *IEEE Trans. Sig. Proc.*, submitted May 1997; see also *First IEEE Workshop on Sig. Proc. Adv. in Wireless Comm.*, pp. 385–388, Paris, France, April 1997.

6. A. W. Brown and T. J. Tukey, "Some distributions of sample means," *Annals of Math. Stat.*, Vol. 17, No. 1, pp. 1–12, 1946.

7. J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. and Mach. Intell.*, Vol. PAMI-8, No. 6, pp. 679–698, November 1986.

8. Z. Cvetkovic and M. Vetterli, "Discrete-time wavelet extreme representation: design and consistent reconstruction," *IEEE Trans. on Sig. Proc.*, Vol. 43, No. 3, pp. 681–693, March 1995.

9. A. Denjean and F. Castanie, "Mean value jump detection: A survey of conventional and wavelet based methods," in *Wavelets: Theory, Algorithms, and Applications*, C. Chui, L. Montefusco, and L. Puccio, eds. (Academic Press, 1994).

10. M. Ghogho and A. Swami, "Fast Computation of the Exact FIM for Deterministic Signals in Colored Noise," submitted to *IEEE Trans. on Sig. Proc.*, June 1997, revised February 1998.

11. M. Gokmen and C. C. Li, "Edge detection and surface reconstruction using refined regularization," *IEEE Trans. Pattern Anal. and Mach. Intell.*, Vol. 15, No. 5, pp. 492–499, 1993.

12. M. Gokmen and C. C. Li, "Multiscale edge detection using first order R-filter," *11th IAPR Intl. Conf. on Patt. Recog., Conf. C: Image, Speech and Sig. Anal.*, pp. 307–310, 1992.

13. A. K. Jain, *Fundamentals of Digital Image Processing* (Prentice-Hall, 1989).

14. R. Kakarala and A. O. Hero, "On achievable accuracy in edge localization," *IEEE Trans. Pattern Anal. and Mach. Intell.*, Vol. 14, No. 7, pp. 777–781, July 1992.

15. J. Koplowitz and V. Greco, "On the edge location error for local maximum and zero-crossing edge detectors," *IEEE Trans. Pattern Anal. and Mach. Intell.*, Vol. 16, No. 12, pp. 1207–1212, 1994.

16. R. Kozick, R. Blum, and B. M. Sadler, "Signal processing in non-Gaussian noise using mixture distributions and the EM algorithm," *31st Asilomar Conf. on Sig., Syst., and Comp.*, Pacific Grove, California, November 1997.

17. R. J. Kozick, B. M. Sadler, and R. S. Blum, "Array processing in impulsive noise: a nonlinear beamformer based on the EM algorithm," *IEEE Trans. Sig. Proc.*, submitted 1998.

18. C. Li, C. Zheng, and C. Tai, "Detection of ECG characteristic points using wavelet transforms," *IEEE Trans. Biomedical Engineering*, Vol. 42, No. 1, pp. 21–28, 1995.

19. L. A. Liporace, "Maximum likelihood estimation for multivariate observations of Markov sources," *IEEE Trans. Info. Theory*, Vol. 28, No. 5, pp. 729–734, 1982.

20. L. Ljung, *System Identification: Theory for the User* (Prentice-Hall, 1987).

21. S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Info. Theory*, Vol. 38, pp. 617–643, 1992.

22. S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Anal. and Mach. Intell.*, Vol. 14, No. 7, pp. 710–732, July 1992.

23. G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions* (Wiley, 1997).

24. K. S. Miller, *Multidimensional Gaussian Distributions* (Wiley, 1964).

25. A. M. Reza and M. Doroodchi, "Cramer-Rao lower bound on locations of sudden changes in a steplike signal," *IEEE Trans. Sig. Proc.*, Vol. 44, No. 10, pp. 2551–2556, 1996.

26. A. Rosenfeld, "A nonlinear edge detection technique," *IEEE Proc.*, pp. 814–816, May 1970.

27. A. Rosenfeld, Y. H. Lee, and R. B. Thomas, "Edge and curve detection for texture discrimination," in *Picture Processing and Psychopictorics*, pp. 381–393 (Academic Press, 1970).

28. A. Rosenfeld and M. Thurston, "Edge and curve detection for visual scene analysis," *IEEE Trans. on Comp.*, Vol. 20, No. 5, pp. 562–569, May 1971.

29. B. M. Sadler, T. Pham, and L. C. Sadler, "Optimal and wavelet-based shockwave detection and estimation," *J. Acoust. Soc. Amer.*, to appear; see also *Proc. Intl. Conf. Acoust., Speech, and Sig. Proc. (ICASSP97)*, Vol. 3, pp. 1889–1992, Munich, Germany, April 1997.

30. A. N. Shiryaev, "The problem of the most rapid detection of a disturbance in a stationary process," *Soviet Math. Dokl.*, No. 2, pp. 795–799, 1961.

31. A. Swami, "Cramer-Rao bounds for deterministic signals in additive and multiplicative noise," *Sig. Proc.*, Vol. 53, pp. 231–244, 1996.

32. H. D. Tagare and R.J.P. deFigueiredo, "On the localization performance measure and optimal edge detection," *IEEE Trans. Pattern Anal. and Mach. Intell.*, Vol. 12, No. 12, pp. 1186–1190, 1990.

33. A. P. Witkin, "Scale-space filtering," *Intl. Joint Conf. on AI (IJCAI-83)*, pp. 1019–1022, Karlsruhe, West Germany, 1983.

34. Y. Xu, J. Weaver, D. Healy, and J. Lu, "Wavelet transform domain filters: a spatially selective noise filtration technique," *IEEE Trans. on Image. Proc.*, Vol. 3, No. 6, pp. 747–758, November 1994.

Appendix A. Cramer-Rao Bounds and M-File Listings

We derive a closed-form expression for the Fisher information matrix (FIM) for a general step-change problem, and give specific results for a sigmoidal (and equivalent tanh) step-change signal model. This leads to a simple expression for the Cramer-Rao bound (CRB) for step change location estimation. This generalizes the discrete-time results of Reza and Doroodchi [25]. The continuous-time model, considered by Kakarala and Hero [14], can also be generalized to an arbitrary signal shape along the lines below.

A.1 General Formulation

The step-change signal is parameterized by a location parameter $\rho = T(n_o + \tau)$, where T is the sampling interval, n_o is an integer, and $\tau \in [0, 1)$ models sample quantization. Initially, we will assume that ρ is nonrandom. Since our discrete wavelet transform (DWT)-based (and other) algorithms yield estimates of n_o, not of ρ, we will later model τ as being random and uniformly distributed over $[0, 1)$. The bound obtained by averaging the FIM for ρ wrt τ may be loosely considered as a bound on the estimate of n_o; it is a bound on the average performance. Note, however, that since n_o is integer valued, its CRB does not formally exist.

Let α denote a parameter that characterizes the step risetime. The noisy signal is given by

$$y(n) = x(n) + v(n) = s(\alpha(n - n_o - \tau)) + v(n), \quad n = 0, ..., N-1\ , \quad \text{(A-1)}$$

where $s(\cdot)$ is the deterministic signal, and $v(n)$ is the additive noise. Note that the observed process $y(n)$ is nonstationary—it has a time-varying mean given by the deterministic signal $x(n)$; its cumulants are identical to those of $v(n)$, so that if $v(n)$ has time-invariant cumulants, so will $y(n)$. In the sequel, we assume that $v(n)$ is iid, zero-mean, non-Gaussian, with finite variance σ_v^2, and pdf $p_v(v)$. The colored noise case can be handled along the lines of Ghogho and Swami [10], and Swami [31].

We will assume that (1) the deterministic signal, $s(\cdot)$, is parameterized by a finite-dimensional parameter vector θ_s, and the noise pdf $p_v(v)$ by a finite-dimensional parameter vector, θ_v; (2) the elements of θ_s do not depend upon those of θ_v; and (3) the noise pdf, $p_v(v)$, satisfies the regularity conditions for the existence of the CRB.

Under these conditions, it was shown by Ghogho and Swami [10] that the FIM for the parameter vector $\theta = [\theta_s, \theta_v]$ is block diagonal. Consequently, the achievable accuracy in estimating θ_s is the same, regardless of whether θ_v is known. This result is well known in the white Gaussian case (less so in the colored Gaussian case); this decoupling between the signal and noise parameters is intuitive, since the signal parameters are related to the mean, and the noise parameters to the cumulants.* Hence, we may assume, without loss of generality, that θ_v is known. In this case, we can use the results of Swami [31] to write the (m, n) element of the FIM corresponding to θ_s as

$$J_{mn} = \left[\frac{\gamma_{v0}}{\sigma_v^2}\right] \left[\sum_{k=0}^{N-1} \frac{\partial x(k)}{\partial \theta_s(m)} \frac{\partial x(k)}{\partial \theta_s(n)}\right], \tag{A-2}$$

where

$$\gamma_{v0} \triangleq \sigma_v^2 \int \left[\frac{dp_v(u)}{du}\right]^2 \frac{1}{p_v(u)}\, du = \sigma_v^2 I_p\,, \tag{A-3}$$

where I_p is the Fisher information for the location parameter of the pdf $p_v(v)$. Thus, the FIM can be written as the product of three terms: the SNR, a pdf-dependent term, and a signal-related term. In the class of symmetric pdfs, $\gamma_{v0} \geq 1$, with equality being attained in the Gaussian case.

As we noted earlier, our DWT-based (and other) algorithms yield integer estimates of n_o; one could, of course, augment this with an interpolation scheme to estimate ρ. Alternatively, one could model τ as being random and take the expectation of the FIM in equation (A-2) wrt τ, and thus obtain an expression for the achievable accuracy in estimating ρ. We will assume that τ is independent of $v(n)$.

Since the pdf of τ is not related to n_o or to the other parameters of $s(\cdot)$, the modified FIM is obtained as

$$J_{mn} = \left[\frac{\gamma_{v0}}{\sigma_v^2}\right] E_\tau \left[\sum_{k=0}^{N-1} \frac{\partial x(k)}{\partial \theta_s(m)} \frac{\partial x(k)}{\partial \theta_s(n)}\right], \tag{A-4}$$

where the expectation is with respect to the pdf of the nuisance parameter τ. Consequently, the modified CRB is given by

*For any iid process with finite moments, the sample estimate of the mean is uncorrelated with that of the variance; further, the two are asymptotically jointly normal. The former is true asymptotically for any weakly mixing process. Further, if we write $x(t)$ as $Ac(t)$, then the achievable bound on A ("amplitude") is decoupled from those on the parameters describing $c(t)$ ("shape").

$$\text{var}\{\hat{\theta}_s\} \geq J^{-1}, \tag{A-5}$$

which may be easily calculated for any specific $s(u)$.

To proceed further, we need an explicit model for $x(n)$ as a function of n_o and the pdf $p(\tau)$. We will assume a uniform pdf for τ (which is, perhaps, the most reasonable pdf for τ), and evaluate the FIM for the general signal case.

A.2 Uniformly Distributed τ

We will focus on the FIM for the location parameter ρ. Because τ is uniformly distributed in $[0, 1)$, we have

$$\begin{aligned} J &= \frac{\gamma_{v0}}{\sigma_v^2} \sum_{n=0}^{N-1} \int_0^1 \left[\frac{\partial x(n)}{\partial \rho}\right]^2 d\tau \\ &= \frac{\gamma_{v0}}{\sigma_v^2} \sum_{n=0}^{N-1} \alpha^2 \int_0^1 [s'(\alpha(n - n_0 - \tau))]^2 d\tau \\ &= \frac{\gamma_{v0}}{\sigma_v^2} \sum_{n=0}^{N-1} \alpha \int_{\alpha(n-n_0-1)}^{\alpha(n-n_0)} [s'(u)]^2 du \\ &= \alpha \frac{\gamma_{v0}}{\sigma_v^2} \sum_{n=0}^{N-1} [G(\alpha(n - n_0)) - G(\alpha(n - n_0 - 1))], \end{aligned} \tag{A-6}$$

where

$$G(u) = \int [s'(u)]^2 du. \tag{A-7}$$

Thus we obtain a *closed-form* expression for the FIM given by

$$J = \alpha \frac{\gamma_{v0}}{\sigma_v^2} [G(\alpha(N - n_0 - 1)) - G(\alpha(-n_0 - 1))]. \tag{A-8}$$

Finally, we may evaluate the CRB via equation (A-5). Note that the FIM is proportional to the risetime parameter α and inversely proportional to the noise variance, as may be expected.

For any reasonable step-approximator, $s(u)$, its derivative will vanish outside a more or less small interval centered around n_0, and we expect that $G(u)$ will attain its asymptotic values rather quickly. The asymptotic value is obtained by assuming that N is large and that n_0 is not too close to either end of the observation window,

$$J_\infty = \alpha \frac{\gamma_{v0}}{\sigma_v^2} [G(+\infty) - G(-\infty)] \,. \tag{A-9}$$

One could consider the signal $s(t)$ as a smoothing filter for the underlying ideal step function; one could then try to find the $s(t)$ that maximizes the FIM given by J_∞, subject to a sensible constraint such as unit energy. It would be interesting to consider the joint estimation of all the signal parameters (α, ρ, and A).

In the case of Gaussian noise, we note that equation (A-8) is also the FIM for the estimate of the nonrandom parameter ρ, when the observations are in continuous time, i.e., $y(t) = s(\alpha(t-\rho)) + v(t)$, $0 \le t \le (N-1)T$, where $v(t)$ is band-limited white Gaussian noise with spectral density σ_v^2. This can be established by mimicking the development in Kakarala and Hero [14].

A.3 Sigmoidal Edge Model

Consider the sigmoidal step change signal described in equation (35),

$$x_s(n) = \frac{m_2 + m_1 e^{-\alpha T(n-n_0-\tau)}}{1 + e^{-\alpha T(n-n_0-\tau)}}. \tag{A-10}$$

Let $A = m_2 - m_1$ denote the step change.

For the nonrandom parameter ρ we obtain

$$J = \frac{(\alpha T)^2 \gamma_{v0}}{16} \frac{A^2}{\sigma_v^2} \sum_{k=0}^{N-1} \cosh^{-4}\left(\alpha T(k - n_0 - \tau)/2\right), \tag{A-11}$$

whose value depends strongly upon τ, but not so strongly upon n_o (unless n_o is close to 0 or $N-1$).

We will now randomize over τ. Using equation (A-10) we obtain

$$\begin{aligned} s(u) &\triangleq \frac{m_2 + m_1 e^{-u}}{1 + e^{-u}} \\ s'(u) &= A \cosh^{-2}(u/2)/4 \\ G(u) &= \frac{A^2}{16} \int [s'(u)]^2 du \\ &= \frac{A^2}{8} [\tanh(u/2) - \frac{1}{3}\tanh^3(u/2)], \end{aligned} \tag{A-12}$$

which via equation (A-8) leads to a closed-form expression for the FIM (which was expressed only as a summation in eq (10) of Reza and Doroodchi [25]),

$$\begin{aligned} J = \frac{(\alpha T) S \gamma_{v0}}{8} &[\tanh(\alpha T(N - n_0 - 1)/2) - \tanh(\alpha T(-n_0 - 1)/2) \\ &- \frac{1}{3}\tanh^3(\alpha T(N - n_0 - 1)/2) + \frac{1}{3}\tanh^3(\alpha T(-n_0 - 1)/2)\Big], \end{aligned} \tag{A-13}$$

where $S \triangleq (m_2 - m_1)^2/\sigma_v^2 = A^2/\sigma_v^2$ is the SNR.

The $\tanh(u)$ function varies from -1 for $u << 0$ through 0 at $u = 0$ to $+1$ for $u >> 1$. Thus, the value of the AFIM for $x_s(n)$ is

$$J_\infty = \frac{\alpha T S \gamma_{v0}}{8} \frac{4}{3} = \frac{\alpha T S \gamma_{v0}}{6}. \tag{A-14}$$

The AFIM is proportional to the SNR and to the risetime parameter αT. The AFIM approximates the FIM extremely well when $N\alpha T >> 1$ (usually a few samples for moderate values of αT).

By re-indexing the observation window from $0, ..., N-1$ to $1, ..., N$ we can write the alternative expression

$$J = \frac{(\alpha T) S \gamma_{v0}}{8} [f(\alpha T N(1-d)/2) - f(-\alpha T d/2)], \tag{A-15}$$

where $d = n_o/N$ is the fractional index of the location parameter, and $f(u) = \tanh(u) - \tanh^3(u)/3$.

A.4 M-File Listings

In this section, we provide listings of Matlab M-files for the forward and inverse Mallat-Zhong discrete wavelet transform (MZ-DWT) algorithm [22, app B].

```
% [WT,S] = fwt(J,x)   Forward wavelet transform
%
%  J = # of scales
%  x = data vector
%  WT = matrix with wavelet transform
%  S = coarse low-pass time series remaining

function [WT,S] = mal_fwt(J, x)

N = length(x);
M = 2*N;

% normalization coefficients

lambda = [1.5, 1.12, 1.03, 1.01];
if J > 4
   lambda = [lambda,ones(1,J-4)];
end
```

```
% filter coefficients

H = [0.125, 0.375, 0.375, 0.125];
G = -1*[-2.0, 2.0];

% convolution offsets

Gn = 2;
for j = 1:J-1
  znum = 2^j - 1;
  Gn = [Gn,((znum+1)/2)+1];
end
Hn = 3;
for j = 1:J-1
  znum = 2^j - 1;
  Hn = [Hn,((znum+1)/2)+znum+2];
end

% compute the WT at each scale
% signal is odd-symmetric periodically extended for borders

S = [fliplr(x),x,fliplr(x)];
WT = [];
figure(1)
for j = 0:J-1

   znum = 2^j - 1; % # of zeros
   Gz = in_zeros(G,znum); % insert zeros into G
   Hz = in_zeros(H,znum); % insert zeros into H

% compute wavelet transform at scale j and store in WT

   Wf = (1/lambda(j+1))*conv(S,Gz); % size(Wf)
   Wf = Wf(N+Gn(j+1):2*N+Gn(j+1)-1);
   WT = [WT, Wf'];

% compute next time series

   S2 = conv(S,Hz);
   S2 = S2(N+Hn(j+1):2*N+Hn(j+1)-1);
   S = [fliplr(S2),S2,fliplr(S2)];
```

```
end

S = S(N+1:2*N);

return


% x = iwt(WT,S)   Inverse wavelet transform
%
%  WT = matrix with wavelet transform
%  S = leftover low-pass time series
%  x = reconstructed time series
%
function x = mal_iwt(WT,S)

[N,J] = size(WT); % J=# of scales, N=data length

% normalization coefficients

lambda = [1.5, 1.12, 1.03, 1.01];
if J > 4
   lambda = [lambda,ones(1,J-4)];
end

% filter coefficients

H = [0.125, 0.375, 0.375, 0.125];
K = [0.0078125, 0.054685, 0.171875];
K = [K, -1*fliplr(K)];

% convolution offsets

Kn = 3;
for j = 1:J-1
  znum = 2^j - 1;
  Kn = [Kn,((znum+1)/2)+2*znum+3];
end
Hn = 2;
for j = 1:J-1
  znum = 2^j - 1;
  Hn = [Hn,((znum+1)/2)+znum+2];
end
```

```
% recursively compute the inverse WT, proceeding down in
% scales, signal is odd-symmetric periodically extended

S=S(:);
S1 = [fliplr(S),S,fliplr(S)]; S1=S1(:);
for j = J:-1:1

   znum = 2^(j-1) - 1; % # of zeros
   Kz = in_zeros(K,znum); % insert zeros into K
   Hz = in_zeros(H,znum); % insert zeros into H

   WTj = WT(:,j); WTj=WTj(:);
   WT_ext = [fliplr(WTj),WTj,fliplr(WTj)]; WT_ext=WT_ext(:);
   A1 = lambda(j)*conv(Kz,WT_ext);
   A1 = A1(N+Kn(j):2*N+Kn(j)-1);
   A2 = conv(Hz,S1);
   A2 = A2(N+Hn(j):2*N+Hn(j)-1);

   S1 = A1 + A2;
   S1 = [fliplr(S1)',S1',fliplr(S1)']; S1=S1(:);

end % end IWT loop

x = S1(N+1:2*N)';

return


% y = ins_zeros(x,n)
%
%   Insert n zeros between elements of vector x
%   and return in y. Used in discrete wavelet transform.
%
function y = in_zeros(x,n)

if n == 0
   y = x;
   return
end

if n>0
   newlen = (n + 1)*length(x); % length of y
```

```
   y = zeros(1,newlen); % new filter vector
   index = 1:n+1:newlen-n;  % indices of data
   y(index) = x; % insert data into y
end

return
```

Distribution

Admnstr
Defns Techl Info Ctr
Attn DTIC-OCP
8725 John J Kingman Rd Ste 0944
FT Belvoir VA 22060-6218

Ofc of the Dir Rsrch and Engrg
Attn R Menz
Pentagon Rm 3E1089
Washington DC 20301-3080

Ofc of the Secy of Defns
Attn ODDRE (R&AT)
Attn ODDRE (R&AT) S Gontarek
The Pentagon
Washington DC 20301-3080

OSD
Attn OUSD(A&T)/ODDDR&E(R) R J Trew
Washington DC 20301-7100

AMCOM MRDEC
Attn AMSMI-RD W C McCorkle
Redstone Arsenal AL 35898-5240

CECOM
Attn PM GPS COL S Young
FT Monmouth NJ 07703

CECOM Sp & Terrestrial Commctn Div
Attn AMSEL-RD-ST-MC-M H Soicher
FT Monmouth NJ 07703-5203

Dir for MANPRINT
Ofc of the Deputy Chief of Staff for Prsnnl
Attn J Hiller
The Pentagon Rm 2C733
Washington DC 20301-0300

Hdqtrs Dept of the Army
Attn DAMO-FDT D Schmidt
400 Army Pentagon Rm 3C514
Washington DC 20301-0460

US Army Edgewood RDEC
Attn SCBRD-TD J Vervier
Aberdeen Proving Ground MD 21010-5423

US Army Info Sys Engrg Cmnd
Attn ASQB-OTD F Jenia
FT Huachuca AZ 85613-5300

US Army Natick RDECActing Techl Dir
Attn SSCNC-T P Brandler
Natick MA 01760-5002

US Army Rsrch Ofc
Attn G Iafrate
4300 S Miami Blvd
Research Triangle Park NC 27709

US Army Simulation, Train, & Instrmntn Cmnd
Attn J Stahl
12350 Research Parkway
Orlando FL 32826-3726

US Army Tank-Automtv & Armaments Cmnd
Attn AMSTA-AR-TD C Spinelli
Bldg 1
Picatinny Arsenal NJ 07806-5000

US Army Tank-Automtv Cmnd Rsrch, Dev, & Engrg Ctr
Attn AMSTA-TA J Chapin
Warren MI 48397-5000

US Army Test & Eval Cmnd
Attn R G Pollard III
Aberdeen Proving Ground MD 21005-5055

US Army Train & Doctrine Cmnd
Battle Lab Integration & Techl Dirctrt
Attn ATCD-B J A Klevecz
FT Monroe VA 23651-5850

US Military Academy
Dept of Mathematical Sci
Attn MAJ D Engen
West Point NY 10996

Nav Surface Warfare Ctr
Attn Code B07 J Pennella
17320 Dahlgren Rd Bldg 1470 Rm 1101
Dahlgren VA 22448-5100

GPS Joint Prog Ofc Dir
Attn COL J Clay
2435 Vela Way Ste 1613
Los Angeles AFB CA 90245-5500

Distribution (cont'd)

DARPA
Attn B Kaspar
3701 N Fairfax Dr
Arlington VA 22203-1714

University of Texas ARL Electromag Group
Attn Campus Mail Code F0250 A Tucker
Austin TX 78713-8029

Hicks & Associates, Inc
Attn G Singley III
1710 Goodrich Dr Ste 1300
McLean VA 22102

US Army Rsrch Lab
Attn AMSRL-CI-LL Techl Lib (3 copies)
Attn AMSRL-CS-AL-TA Mail & Records Mgmt
Attn AMSRL-CS-EA-TP Techl Pub (3 copies)
Attn AMSRL-IS-TA B Sadler (50 copies)
2800 Powder Mill Rd
Adelphi MD 20783-1197

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1998	Final

4. TITLE AND SUBTITLE
Analysis of Wavelet Transform Multiscale Products for Step Detection and Estimation

5. FUNDING NUMBERS
PE: 61102A

6. AUTHOR(S)
Brian M. Sadler and Ananthram Swami

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
Attn: AMSRL-IS-TA (email: bsadler@arl.mil)
2800 Powder Mill Road
Adelphi, MD 20783-1197

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-1664

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
2800 Powder Mill Road
Adelphi, MD 20783-1197

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
AMS code: 611102H4411
ARL PR: 8FE310

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

We consider discrete wavelet transform (DWT) multiscale products for detection and estimation of steps. Here the DWT is an overcomplete approximation to smoothed gradient estimation, with smoothing varied over dyadic scale, as developed by Mallat and Zhong. We show that the multiscale product approach, as first proposed by Rosenfeld for edge detection, is a nonlinear whitening transformation. We characterize the resulting non-Gaussian heavy-tailed densities. The results may be applied to edge detection with a false alarm constraint. The response to impulses, steps, and pulses is also characterized. A general closed-form expression for the Cramer-Rao bound (CRB) for discrete and continuous-time step-change location estimation in independent identically distributed non-Gaussian noise is developed, generalizing previous results. We consider location estimation using multiscale products, and compare results to the appropriate CRB.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Wavelets, edge detection, multiscale, estimation	51
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102